WHAT IS CLAIMED IS:

1. A method of storing page link information comprising:

obtaining page link information for a set of pages, the page link information including for each page in the set a row of page identifiers of other pages;

arranging the rows of page identifiers in a particular order;

for each respective row:

identifying a reference row, if any, that best matches the respective row in accordance with predefined row match criteria; and

encoding the respective row as an identifier for the identified reference row, if any, a set of deletes representing page identifiers in the identified reference row not in the respective row, and a set of adds representing page identifiers in the respective row not in the identified reference row.

2. The method of claim 1, wherein the encoding for each respective row includes Huffman coding values representing the set of deletes and the set of adds for each respective row.

3. The method of claim 1, wherein the encoding for each respective row includes delta encoding the set of deletes and delta encoding the set of adds for each respective row.

4. The method of claim 1, wherein the encoding for each respective row includes

delta encoding the set of deletes and delta encoding the set of adds for each respective row; and

Huffman coding the delta encoded set of deletes and delta encoded set of adds for each respective row.

5. The method of claim 4, including

sorting the page identifiers in each row in numerical order prior to performing the encoding.

6. The method of claim 5, wherein the encoding includes generating a row distance value that identifies the identified reference row and Huffman coding the row distance value.

7. The method of claim 4, including

when no reference row exists for a respective row, encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row, delta encoding the set of adds for the respective row, and Huffman coding the delta encoded set of adds for the respective row.

8. The method of claim 1, including

when no reference row exists for a respective row, encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row.

9. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

a first module for obtaining page link information for a set of pages, the page link information including for each page in the set a row of page identifiers of other pages; and

a second module for storing the page link information, including instructions for:

arranging the rows of page identifiers in a particular order;

for each respective row:

identifying a reference row, if any, that best matches the respective row in accordance with predefined row match criteria; and

encoding the respective row as an identifier for the identified reference row, if any, a set of deletes representing page identifiers in the identified reference row not in the respective row, and a set of adds representing page identifiers in the respective row not in the identifier reference row.

10. The computer program product of claim 9, wherein the encoding instructions of the second module include instructions for Huffman coding values representing the set of deletes and the set of adds for each respective row.

11. The computer program product of claim 9, wherein the second module includes instructions for delta encoding the set of deletes and delta encoding the set of adds for each respective row.

12. The computer program product of claim 9, wherein the encoding instructions of the second module include instructions for delta encoding the set of deletes and delta encoding the set of adds for each respective row, and for Huffman coding the delta encoded set of deletes and delta encoded set of adds for each respective row.

13. The computer program product of claim 12, wherein the second module includes instructions for sorting the page identifiers in each row in numerical order prior to performing the encoding.

14. The computer program product of claim 13, wherein the encoding instructions of the second module include instructions for generating a row distance value that identifies the identified reference row and Huffman coding the row distance value.

15. The computer program product of claim 12, wherein the second module includes instructions, used when no reference row exists for a respective row, for encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row, delta encoding the set of adds for the respective row, and Huffman coding the delta encoded set of adds for the respective row.

16. The computer program product of claim 9, wherein the second module includes instructions, used when no reference row exists for a respective row, for encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row.

17. A web crawler system, comprising:

a central processing unit for performing computations in accordance with stored procedures;

a network interface for accessing remotely located computers via a network;

memory, coupled to the central processing unit, for storing procedures and data, including:

a web crawler module, executable by the central processing unit, for downloading a set of pages from remotely located servers via the network interface;

a first module for obtaining page link information from the set of pages, the page link information including for each page in the set a row of page identifiers of other pages; and

a second module for storing the page link information, including instructions for:

arranging the rows of page identifiers in a particular order;

for each respective row:

identifying a reference row, if any, that best matches the respective row in accordance with predefined row match criteria; and

encoding the respective row as an identifier for the identified reference row, if any, a set of deletes representing page identifiers in the identified reference row not in the respective row, and a set of adds representing page identifiers in the respective row not in the identified reference row;

18. The system of claim 17, wherein the encoding instructions of the second module include instructions for Huffman coding values representing the set of deletes and the set of adds for each respective row.

19. The system of claim 17, wherein the encoding instructions of the second module include instructions for delta encoding the set of deletes and delta encoding the set of adds for each respective row.

20. The system of claim 17, wherein the encoding instructions of the second module includes instructions for delta encoding the set of deletes and delta encoding the set of adds for each respective row, and for Huffman coding the delta encoded set of deletes and delta encoded set of adds for each respective row.

21. The system of claim 20, wherein the second module includes instructions for sorting the page identifiers in each row in numerical order prior to performing the encoding.

22. The system of claim 21, wherein the encoding instructions of the second module include instructions for generating a row distance value that identifies the identified reference row and Huffman coding the row distance value.

23. The system of claim 20, wherein the second module includes instructions, used when no reference row exists for a respective row, for encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row, delta encoding the set of adds for the respective row, and Huffman coding the delta encoded set of adds for the respective row.

24. The system of claim 17, wherein the second module includes instructions, used when no reference row exists for a respective row, for encoding the respective row by encoding a null reference row identifier and a set of adds representing the page identifiers in the respective row.